Published in final edited form as:
Dev Biol. 2011 September 15; 357(2): 505–517. doi:10.1016/j.ydbio.2011.06.016.

Precise *cis*-regulatory control of spatial and temporal expression of the *alx-1* gene in the skeletogenic lineage of *s. purpuratus*

Sagar Damle and **Eric H. Davidson***
Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

Deployment of the gene regulatory network (GRN) responsible for skeletogenesis in the embryo of the sea urchin *Strongylocentrotus purpuratus* is restricted to the large micromere lineage by a double negative regulatory gate. The gate consists of a GRN subcircuit composed of the *pmar1* and *hesC* genes, which encode repressors and are wired in tandem, plus a set of target regulatory genes under *hesC* control. The skeletogenic cell state is specified initially by micromere-specific expression of these regulatory genes, *viz*. *alx1*, *ets1*, *tbrain* and *tel*, plus the gene encoding the Notch ligand Delta. Here we use a recently developed high throughput methodology for experimental *cis*-regulatory analysis to elucidate the genomic regulatory system controlling *alx1* expression in time and embryonic space. The results entirely confirm the double negative gate control system at the *cis*-regulatory level, including definition of the functional HesC target sites, and add the crucial new information that the drivers of *alx1* expression are initially Ets1, and then Alx1 itself plus Ets1. *Cis*-regulatory analysis demonstrates that these inputs quantitatively account for the magnitude of *alx1* expression. Furthermore, the Alx1 gene product not only performs an auto-regulatory role, promoting a fast rise in *alx1* expression, but also, when at high levels, it behaves as an autorepressor. A synthetic experiment indicates that this behavior is probably due to dimerization. In summary, the results we report provide the sequence level basis for control of *alx1* spatial expression by the double negative gate GRN architecture, and explain the rising, then falling temporal expression profile of the *alx1* gene in terms of its auto-regulatory genetic wiring.

Keywords

Alx1 gene; cis-regulation; skeletogenic micromere lineage; gene regulatory network; tagged reporter assay

Introduction

Developmental gene regulatory networks (GRNs) are models that explain embryonic specification functions in terms of a hierarchical matrix of genomically encoded information processing events. The GRN that encodes pre-gastrular development of the *S. purpuratus* large-micromere/skeletogenic mesenchyme (SM) lineage is to date among the most complete and well studied (Davidson et al., 2002; Oliveri et al., 2002; Oliveri et al., 2003;

*Corresponding author. Fax: +1 818 583 8351. Davidson@caltech.edu..

Oliveri et al., 2008). The specification of the large micromeres is initiated by action of a double-negative regulatory gate, whereby micromere specific expression of the repressor gene *pmar1* in turn represses transcription of the ubiquitously-driven repressor gene, *hesc* (Ettensohn et al., 2003; Gao and Davidson, 2008; Oliveri et al., 2003; Revilla-i-Domingo et al., 2007). This regulatory gate can be shown to operate as a logic processing device (Peter and Davidson, 2009). It is also of evolutionary importance, as it has been considered the focal point in the GRN for the redeployment of the preexisting adult skeletogenic apparatus to the micromere lineage early in the evolutionary divergence of the euchinoids (Gao and Davidson, 2008). Understanding the sequence basis of the mechanism by which HesC repression unlocks the skeletogenic program will illuminate the pathway by which such network co-options may have occurred.

Alx1, ets1, tbrain and *tel* are the earliest transcription factors defining the definitive zygotic skeletogenic micromere (SM) regulatory state, and together with the gene encoding the Notch ligand Delta, these genes are expressed immediately downstream of the double negative regulatory gate. In previous work the sequence basis of HesC repression, and thus in the SM lineage the release from this repression, has been identified at the genomic *cis*-regulatory level for the *tbrain* gene (Wahl et al, 2008) and the *delta* gene (Revilla et al, 2007; Smith et al, 2008). But despite its key importance for the subsequent developmental processes of the SM lineage (Ettensohn et al., 2003), no *cis*-regulatory information has been available for the *alx1* gene. *Alx1* encodes the first invertebrate member of the Cart1/Alx3/Alx4 family of Paired-class homeodomain proteins, also known as ‘Group-I Aristaless-like’ factors (Ettensohn et al., 2003). In addition to its homeodomain, the protein encoded by the *Strongylocentrotus alx1* gene shares with vertebrate CART family members the presence of a charged domain near the N-terminus, an OAR/Aristaless domain at the C-terminus, and a generally proline-rich primary sequence. Alx transcription factors appear to share an ancient, conserved role in skeletogenic development. The *alx1* gene is expressed in both juvenile sea urchin and sea star skeletonization centers (Gao and Davidson, 2008), as well as in the sea urchin embryo, while several Group-I *aristalless* like genes are expressed during vertebrate embryogenesis in the mesenchymal cells that form the craniofacial and appendicular skeleton (Beverdam and Meijlink, 2001; Qu et al., 1997). Loss-of-function mutations in these genes lead to defects in skeletal elements in mice. Though the downstream effector molecules for skeletogenesis in echinoderms and vertebrates are different (the biominerals per se are non-homologous), these similar expression patterns may reflect functional conservation of regulatory cassettes controlling skeletogenic state specification in the ancestral deuterostome.

Alx1 is first expressed in the large daughters of 4th cleavage micromeres and its spatial expression is restricted to the descendants of this cell lineage, the skeletal mesenchyme (SM), for the remainder of embryogenesis. The quantitative temporal profile of *alx1* expression pattern is fairly complex, as was first observed by P. Oliveri (unpublished), and illustrated here in Fig.1. Expression begins around 7.5 hours post fertilization (hpf), and peaks twice during embryogenesis, at first sharply at pre-hatching blastula stage (10-12 hpf) and then more gradually at mesenchyme blastula stage (23-25hpf). Previous studies based on morpholino antisense interference have suggested that *alx1* is driven in SM cells by the Ets1 transcription factor (Ettensohn et al., 2003; Sharma and Ettensohn, 2010).

GRN models can ultimately be validated by *cis*-regulatory analysis, in which the predicted target sites are identified and their predicted functionalities demonstrated. This level of structure/function analysis immediately identifies the genomic regulatory code, the functional meaning of which is explicitly predicted in the GRN model, and *cis*-regulatory analysis is also the final arbiter of direct vs. indirect genetic interactions. Here we deconstruct the *alx1* expression pattern during early development into activation and

repression components, by identifying the genomic regulatory sequences responsible for these functions. Thus we have experimentally identified, and by mutation functionally characterized the genomic target sites responsible for direct spatial repression by the *hesc* gene product, and for activation by Ets1. In addition, we demonstrate that Alx1 protein is both an immediate, direct, auto-activator, and at higher concentrations an auto-repressor, and reveal the biochemical and gene-regulatory network architectural features that permit these opposing roles. These regulatory interactions provide an explanation for both the lineage-specific spatial expression of the *alx1* gene and its kinetic expression profile.

Materials and Methods

Injection and scoring of reporter constructs

Sea urchin eggs and sperm were isolated and prepared for injection as described (Cheers and Ettensohn, 2004). Small constructs were injected in a solution containing 120mM KCl, and 30ng/ul of carrier DNA. BAC-GFP constructs were injected without carrier DNA. In barcoded GFP reporter experiments, multiple DNA constructs were mixed and co-injected at a total concentration of 0.9ng/ul (roughly 110 total copies per 2pl), and injection volume per egg was approximately 10-20pl. mRNA constructs were injected without carrier DNA at concentrations ranging from 1ug/ul to 100ng/ul.

Isolation of Alx1 BAC and phylogenetic footprinting

An *alx1* BAC (Sp_BAC_042I08_L) was isolated from a *Strongylocentrotus purpuratus* BAC library as described (Lee et al., 2007), using a partial cDNA probe. The BAC was mapped to get an estimate of the minimum distance between the *alx1* coding sequence and the termini of the insert. Mapping was performed by digesting the BAC with Kpn1 and gel-purifying the individual restriction fragments. These fragments were used as templates for QPCR. Each fragment was assayed for the presence of vector sequence (pBACe 3.6) and for *alx1* exon sequences. The mapping step was used to preclude BACs that are not desirable for *cis*-regulatory analysis because individual restriction fragments contain both vector sequence and *alx1* coding sequence, i.e., in which the *alx1* gene borders the edge of the BAC insert.. Similar procedures were also used to isolate a *Lytechinus varieagatus alx1* BAC (Lv_BAC_007J11_L).

Phylogenetic footprinting between the *S.p.* and *L.v. alx1* BAC sequences was performed using SeqComp, and visualized by the Family Relations software package (Brown et al., 2002). Seqcomp was performed using a 50bp window and 80% sequence similarity.

Generation of BAC GFP reporter and deletion constructs by homologous in vitro *recombination*

The parental BAC is here referred to as *alx1:*GFP BAC. It was generated by homologous recombination as described (Court et al., 2002). The targeting cassette contained the GFP coding sequence, an SV40 poly-adenylation site, and the kanamycin gene, flanked by flp-recombinase target sites. The targeting cassette was amplified using primers with 5′ tails homologous to the insertion site as follows (*alx1*-specific targetting sequence underlined):

Alx-GFP-cassette_Forward:
GCCTTTTCTTAGGATTTTGTCGTGCCGAGACTTTACTCAATATTGATGAGCAAGGGCGAGGAACT

Alx-GFP-cassette_Reverse:
AGTTTACTTACACGTCGCTAAGCACGGCATTGAGGGGTAAAACAATCGAAGAGCTATTCCAGAAGTAGTGA

Homologous recombination with this cassette replaced the first 36bp of coding sequence from the 3′ portion of *alx1* exon1 with the GFP cassette. After recombination, the kanamycin resistance gene and bacterial regulatory DNA were removed by induction of the *flippase* gene, leaving a 126 bp artifact containing one 45 bp flp-recombinase site downstream of the SV40 3′UTR.

The presence of an extraneous flp site acts as an anchor point for subsequent homologous recombination experiments using the flp-recombinase. Therefore, an alternative strategy involving Galk positive/negative selection (Warming et al., 2005) was used to generate mutational variants of the *alx1:*GFP BAC. A targeting cassette containing galK was amplified using tailed primers containing sequences flanking two putative HesC binding sites as follows (*alx1*-specific targeting sequence underlined):

HesC flanking site Forward:
ACTCTTGACCAATGACCGTGCCCGAAGCCCAGCGGTGTATAATAGCCTGTTGACAATTAATCATC

HesC flanking site Reverse:
GAGCGAGAGTGAAAATCGGCGAGTGCTTCGGCGGAGCGAAGAAACTCAGCACTGTCCTGCTCCTT

Recombinant Galk-containig BACs were screened for proper insertion using primer pairs that bridged the insertion site, and later confirmed by sequencing. A second homologous recombination was performed using a cassette containing the desired mutated sequence and flanked by 150bp homologous target-sequence. Recombinants were isolated by negative selection for Galk as described.

Generation of cis-regulatory reporter constructs for GFP scoring and QPCR

Cis-regulatory reporter constructs were generated by fusing a putative regulatory sequence to *alx1* basal promoter and to a GFP cassette in two successive fusion PCR steps as described (Hobert, 2002). An adaptamer with the following sequence was used to fuse putative *cis*-regulatory modules to the Alx1 basal promoter:
AGCTTGATATCGAAGTCCTGCAG

The set of 13 “barcoded” GFP vectors that we developed for high throughput *cis*-regulatory analysis (Nam et al., 2010) were individually fused to various regulatory DNA/promoter construct combinations, mixed into the same injection solution, and injected in fertilized eggs. The GFP “barcode” sequence tags are detected independently using specific QPCR primers (Nam et al, 2010). QPCR was used to measure reporter activity of each tag GFP construct quantitatively, and the results were normalized to the number of integrated genomic copies of that tag as described (Revilla-i-Domingo et al., 2004). GFP expression as measured by the abundance of unique tags was then also normalized for minor tag-specific differences in transcript half-life. This was done by assaying the variability of expression of 13 tag reporters when driven by *identical* active cis-regulatory modules. This measurement was repeated 5 times and used as a normalization standard in all tag experiments (Supplemental Figure1). A negative control was constructed by fusing the basal promoter-GFP construct to a series of nonfunctional genomic fragments ~500-1000 bp in length. Expression from this construct is used to set a baseline for all expression data (Supplemental Figure2)

Mutation of putative transcription factor binding sites within reporter constructs

Site-specific mutation of reporter constructs was performed using fusion PCR with primers overlapping the target sequence but containing the desired mutation or deletion. Each primer

was approximately 45bp long and included the target site disruption and 20 bp of unmutated flanking sequence. Primer sequences used to generate the following mutants are described in Supplemental Table 1: et1s (x5), hesc proximal, hesc distal, tcf proximal, tcf distal, alx distal (x3), alx-proximal.

Overexpression of mRNA encoding monomeric and tethered obligate dimer forms of Alx1

Monomeric *alx1* mRNA was constructed by amplifying the full length coding sequence from a population of 11.5 hpf cDNAs. An obligate dimer of *alx1* was generated through fusion PCR by joining two copies of the *alx1* coding sequence with coding sequence for a glysine-serine tether (GGGGS)x3 kindly provided by Joshua Klein, Caltech. Each construct was cloned into p-gemT vector and capped mRNAs were synthesized using the T7 mMessage mMachine RNA Transcription Kit (Ambion) and polyadenylated using the poly-A synthesis kit (Ambion). These synthetic mRNAs were injected into fertilized eggs as described above.

Results

Structure of the alx1 genomic locus

The *alx1* locus can be found on NCBI genomic scaffold NW_001306657. In addition to *alx1*, this scaffold contains three other genes: *alx-related1*, *LOC583266* and *pit54-related* (Figure2A). A comparison of scaffold sequence to *alx1 cDNA* (NM_214644, GLEAN3_25302) revealed that the *alx1* gene contains 6 exons extending over 36kb of genomic DNA: a 251bp 5′UTR is contained within exon1 and a 3360bp 3′UTR is in exon 6. The Alx1 homeodomain coding region is encoded in exons 2-4. A putative transcriptional start site, containing a canonical TATA box was identified 49bp upstream of the 5′UTR (figure 2A).

Recapitulation of endogenous alx1 expression by an alx1 GFP BAC reporter

A 129kb BAC (042I08_L) containing *alx1* was sequenced and found to include all 6 exons as well as 35kb of upstream and 57kb of downstream flanking sequence (Figure 2B). Using bacterial homologous recombination, we inserted a GFP reporter cassette within exon1 at the start of the *alx1* coding sequence (Figure2B). When injected into fertilized eggs and assayed for expression during development, the BAC-GFP reporter activity closely resembled endogenous *alx1* expression from early blastula through to late gastrula stage (Figure 2C, 2D). QPCR of the alx-GFP BAC reporter shows that expression begins at 8hpf and reaches peak activity of roughly 80 copies GFP per construct (cp/construct) at 9-10 hpf. Expression drops to 10 cp/construct at 16-18 hpf and is followed by a second smaller peak of 25 cp/construct at 24 hpf (Figure 2C). For comparison the kinetics of endogenous *alx1* expression from Fig,1 are co-plotted in Fig.2C. The kinetics are slightly different; note that the turnover of *gfp* mRNA is slightly slower than that of *alx1* mRNA, as evident in the falling portions of the respective curves (peak to trough 9h for *gfp* mRNA vs 6h for *alx1* mRNA), but both display the same sharp rise in mRNA level followed by an abrupt decline. In living injected embryos, GFP fluorescence was initially detected in the large micromeres as early as 12hpf (several h. are required for GFP protein to fold into its native fluorogenic form in sea urchin egg cytoplasm at 15°). Expression persisted in the large micromeres and their descendents, the SM lineage, throughout the remaining 72 hours of embryonic development (Figure 2D), and there was minimal ectopic expression (Table1). These observations demonstrate that as would be expected from the position of the gene in the BAC, the *alx1* BAC GFP reporter contains all the necessary *cis*-regulatory control apparatus to recapitulate the correct spatial and temporal expression of *alx1* in the embryo.

Application of high throughput technology for alx1 cis*-regulatory analysis*

In order to accelerate the collection of *cis*-regulatory data, we employed a new high throughput approach for rapid, parallel discovery and quantitative characterization of regulatory DNA sequence (Nam et al., 2010). This method permits the simultaneous introduction of multiple *cis*-regulatory expression constructs into the same batch of eggs. The activity of each individual reporter construct in the injected mixture is subsequently deconvolved by identification and quantification of its transcription product using sequence tags incorporated in the constructs, which act as unique “barcodes”. The activity of each “barcoded” reporter is thus assayed independently in the nucleic acid extracted from the embryos by QPCR. This strategy enabled experimental measurement of the regulatory functions of multiple 0.5-2kb genomic DNA sequences, together with positive and negative controls, in each experiment. Control experiments in which all the construct tag vectors were driven by the same known active *cis*-regulatory module displayed subtle, i.e., less than 2-fold, variation in tag-specific expression (Supplementary Fig.1). While this amount of variation does not affect screening for weak or strong activator modules, it could interfere with quantitative detection of minor differences, for instance in assessing the effects of site mutations. To eliminate this source of variation, a normalization factor was obtained for each tag by averaging tag-specific activity over 5 repeated control experiments. These tag specific normalization factors were applied to all subsequent data obtained with the high throughput tag system.

Scanning for functional non-coding sequence patches near the alx1 gene

To identify putative regulatory modules, we looked for sequence patches in the vicinity of the *alx1* gene that are conserved between *S. purpuratus* and *L. variegatus* genomes. Phylogenetic footprinting was carried out using Family Relations software (Brown et al., 2002) to compare *alx1* BAC sequences from the two species, using a 50 bp sliding window within which >80% sequence identity was required. The region of overlap between the *Sp* and *Lv* BAC sequences was approximately 75 kb, and included all 5 *alx1* introns as well as 35kb of upstream genomic DNA (note that there is very little intergenic space between the 3′ end of the *alx1* gene and the two following downstream genes; Fig.2a). This analysis identified 14 non-coding conserved sequence patches (labeled A-N) that lay within 25kb of exon1 (Figure 3A). The conserved patches ranged in size from 247bp to 1735bp and had an average size of 1kb. These sequences were isolated by PCR and fused into a set of barcode tag vectors that contained the *alx1* basal promoter with the GFP coding sequence serving as the reporter (see Methods). Of all fragments tested (C through N), only fragment J generated levels of expression higher than background at 10hpf (Figure3B), which corresponds to the first peak of *alx1* expression (Fig.1). Fragments A and B, which are not included in Fig.3B, were independently tested and found to be inactive at both 10 and 24hpf. ;

The J’ construct, which contains the J-module and alx basal promoter and endogenous 5′UTR, expresses faithfully in the skeletogenic cell lineage; 89% of expressing embryos produce GFP in skeletogenic cells with very little ectopic expression (Table 1). Kinetics of J’ output measured as GFP mRNA, and examples of J’ construct spatial expression are shown in SFig.4. To exclude the possibility of additional functional regulatory sequences that are non-conserved or only weakly conserved, an additional series of serially truncated BAC sequence fragments was tested. These were directly amplified from the *alx1* GFP BAC. These non-conserved sequences were labeled in lowercase (a-m) such that, for instance, element ‘a’ lies between elements A and B. Figure 4 shows that reporter construct i→J, which includes nonconserved region i and conserved module J, was the shortest reporter capable of quantitatively matching *alx1* BAC GFP output levels per incorporated construct molecule (at 11.5hpf). To be certain that region i contained no short conserved elements, a high-resolution Family Relations analysis with window size of 10bp and 90%

similarity was performed, but no regions of conservation were found (Supplemental Fig.3). Comparison with the activity of J' alone shows that inclusion of non-conserved region i materially boosts the output of J' (though it is not capable of driving expression by itself), but no other additional upstream sequence further affected construct output (Fig.4B).

Direct transcriptional repression of alx1 *by HesC*

As reviewed briefly above, expression of the initial tier of genes constituting the SM regulatory state, including *alx1*, is confined to the SM lineage by HesC repression everywhere else. Expression of these genes is permitted to occur in the micromere descendants because the initial gene of the double negative gate, *pmar1*, specifically prevents *hesC* transcription in these cells, where *pmar1* is activated soon after the lineage founder cells are born at 4th cleavage (Oliveri et al., 2002; 2003; Revilla et al, 2007). When *HesC* MASO is injected into fertilized eggs, *alx1* expression is up-regulated 4-7 fold (Revilla-i-Domingo et al., 2007). Taken together with the short time lag, 2.5-3h, between *pmar1* and *alx1* activation, these results predict that HesC directly represses the *alx1* gene (cf. the kinetic study of gene cascades in this embryo; Bolouri and Davidson, 2003). Functional promoter-proximal Hesc binding sites of identical sequence have already been identified in two other genes that are activated coordinately with *alx1* and are also under control of the double negative gate, *tbrain* and *delta* (Smith and Davidson, 2008; Wahl et al., 2009). Putative HesC sites are also present near the promoter for the *ets1* gene, another double negative gate target gene (S. Damle, unpublished data). Thus we sought to determine if functional HesC binding sites exist in the accurately expressed J' *alx1* minimal expression construct. The *S. purpuratus hesC* gene is a member of the Hairy/E(spl) family of transcription factors (Howard-Ashby et al., 2006), which prefer class-B (CACGTG) or class-C (CACGCG) E-box sites (Fischer and Gessler, 2007). The repressive functions of Hairy/E(spl) members are mediated by their cofactor Groucho, an obligate transcriptional repressor, with which they specifically interact (Grbavec and Stifani, 1996; Paroush et al., 1994). *Hesc* is indeed expressed ubiquitously outside the SM lineage in the pre-hatching embryo, and its expression becomes more intense in the endoderm prior to hatching (Revilla-i-Domingo et al., 2007). A GFP-BAC reporter for the *hesC* gene corroborates this observation, showing stronger expression in presumptive endoderm at 24 hpf compared to ectoderm and NSM (Smith and Davidson, 2008).

Two class-C E-box sites in fact flank the TATA-box of the *alx1* J' -construct. The proximal class-C site sits only 10bp downstream of the TATA box, while the distal site lies 47bp upstream (Fig.5A). When both sites are mutated (Fig.5B), the J' construct produces a relatively enormous amount of ectopic expression in ectoderm, veg1, and veg2 lineages at 18hpf (Fig.5C, left panel). Expression is quantitatively up-regulated 2-fold at 10-12hpf as a result of the mutation of the two HesC sites (Fig.5C, middle panel). Assessed as early as 9.5hpf, only 1.5-2h and in micromere descendants at 7th cleavage (transcripts of *alx1* can first be detected only at 6th cleavage), the expression level of the construct lacking the two HesC sites is 2-4x that of the native J' construct (Fig.5C, right panel). Furthermore, by 8.25h, as early as quantitatively reliable results can be obtained, *hesC* MASO markedly derepresses the wildtype J' construct, an experiment shown in Fig.S9. All of these results conform directly with the network prediction that a ubiquitously expressed activator initiates *alx1* expression and that the interaction of Hesc at the proximal and distal binding sites is responsible for blocking *alx1* expression in all nonskeletogenic cells of the embryo *ab initio*. We complemented this result with a *cis-trans* test, which shows that unlike the native J' -construct, the double-mutant J' construct is largely insensitive to co-injection with *hesc* MASO (Fig.5C, middle panel). Note that in this experiment the wildtype J' construct expresses at higher levels in the presence of the *hesc* MASO than in its absence: this is because the driver is Ets1, as we confirm below, and the *ets1* gene is also derepressed by

hesc MASO. The main point, however, is that the wildtype and the J' construct lacking HesC target sites behave the same in response to *hesc* MASO.

Do these two HesC sites control spatial expression in the context of the whole gene as well as in the minimal J' construct? Perhaps the whole gene regulatory system might include some additional spatial control mechanism, so that although mutation of the HesC sites in J' construct indeed produces ectopic expression, mutation of these two 6bp sequences in the native gene might fail to derange normal spatial expression, which in context could be controlled by other interactions as well. Recalling that the genomic region 3′ of the gene was not included in the conserved sequence scan (Fig.3A), such additional interactions could, for example, be mediated at an unexplored module located in this region, even within or beyond the closely flanking genes. To examine such possibilities, we used BAC recombineering to remove each HesC site from the complete *alx1* GFP BAC, and tested the spatial expression of the mutant BAC constructs. But the results were essentially the same as for the mutated J' construct. In the complete context of the *alx1* GFP BAC, when either the proximal or distal HesC binding site is mutated, striking ectopic expression results, as illustrated in Fig.5D. Statistics collected for embryos injected with the mutant *alx1* BAC GFP reporters showed that, at 24hpf for example, there was 4× more ectopic expression in NSM when the proximal HesC site alone is mutated; and when the distal HesC site is mutated there was 2× more expression in NSM, 5× more expression in the endoderm, and 2.5× more ectopic expression in the ectoderm, than in the parental BAC construct (Table 1). Taken together, these *cis*-regulatory results plus the *hesC* MASO data indicate that Hesc is in fact the direct input responsible for repressing *alx1* gene expression outside of the SM lineage from the beginning of its expression up through mesenchyme blastula; that these HesC sites function to restrict expression to the SM lineage in the context of the whole gene; and that the genomic locus of the repressive input is specifically the two class-C, E-boxes included in J' -construct that flank the promoter, both of which are necessary for complete repression. In their absence, expression is not properly restricted. In summary, the data in Figs. 5 and Supplementary figure 9 show that as predicted, HesC repression controls the activity of the *alx1* regulatory system from as early as direct measurements can be made, and in our view is the only effective spatial restriction system active in early development in this gene.

Cis-regulatory identification of the activator of early alx1 *gene expression*

Ets1 has been thought to provide a positive input to *alx1* (Ettensohn, 2003; Sharma and Ettensohn, 2010; Oliveri et al, 2008), and *ets1* mRNA of maternal origin is initially present at very high levels, 30,000 molecules per egg. The *ets1* gene is expressed zygotically before 12hpf (Rizzo et al., 2006). It is proposed that Ets1 is not the earliest driver of *alx1* expression because injection of mRNA encoding a "dominant negative" Ets protein lacking the transactivation domain fails to prevent *alx1* activation (Sharma and Ettensohn, 2010). However, expression of a dominant negative form of Ets1 might not fully block the function of maternal Ets1 protein. Furthermore, although immunostaining studies indicate that endogenous Ets1 protein does not enter the nuclei of presumptive PMCs until after alx1 has been activated (Sharma and Ettensohn, 2010; Yajima et al., 2010), there might be functional levels of Ets1 protein present in the nuclei of these cells that are too low to be detected by immunostaining. Two putative Ets binding sites of the form ($^{C}/_{A}$GGAA) are present in the J module between −423 and −299, and three additional Ets binding sites in the i' sequence between −1105 and −795 (Fig.6A,B). Mutation of the two J-module Ets sites within the context of the larger i' →J construct decreased expression by over 50%, and deletion of these sites by about 3-fold (Fig.6C). In contrast, mutational analysis of the 3 sites within the more distal region i proved they are not strongly required. An additional more detailed series of deletions summarized in Supplemental Figure 5A displayed no additional cryptic sites in J' .

A further experiment demonstrates that Ets1 specifically interacts with *ets* binding sites identified in i' →J. We measured expression of the construct of the 5-site ets mutant form of i' →J in the presence of *ets1* MASO. If Ets1 is a direct activator of *alx1* then the activity of the 5-fold mutant should be insensitive to *ets1* MASO and should equal the level of expression of normal i' →J in *ets1* MASO embryos, and this was the result obtained (Fig. 6C).

Alx1 modulates its own transcription

In earlier work it was found that *alx1* is apparently repressed by its own gene product, since post-hatching embryos injected with *alx1* MASO displayed elevated levels of *alx1* transcript (Ettensohn et al., 2003) . We carried out similar experiments but assayed the results by QPCR at the three key periods in the *alx1* expression time-course: at the first expression peak, 11.5hpf; at the lowest point, 16hpf; and at the late expression peak, 24hpf (Figure 7A). MASO knockdown led to a 4-fold reduction in *alx1* at the first peak of expression, which was followed by an equal but opposite fold increase in expression after 16hpf. These later results were consistent with the earlier post-hatching blastula measurements (Ettensohn et al., 2003), but the early time point in Fig.7a revealed that *alx1* also has a positive auto-regulatory input on its own expression during the dramatic early rise in transcript level. Because the i' →J construct quantitatively recapitulates the early *alx1* expression profile (Fig.4B), it must contain the regulatory sequences responsible for both auto-activation and repression.

Homeodomain transcription factors bind to regulatory DNA via a helix-turn-helix motif that canonically recognizes AT-rich binding sites including the element TAAT. Members of the vertebrate class of Cart/Alx family contain a "Q50" Paired-type homeodomain that can dimerize cooperatively, and the dimer binds to a pair of palindromic half-sites that are separated by 3bp, known as P3 sites (Wilson et al., 1993). Five putative monomeric Alx1 binding sites in module i' →J were identified, but mutation of all of these sites collectively had no effect on construct expression levels (at 11.5hpf; data not shown). We then looked for P3 sites of the form TAATNNNATTA. One such P3 sequence exists within the 5′UTR and 3 others in region 'i' (Fig.7C). Mutation of all four of these sites indeed led to a 2.5-fold drop in i→J activity at the 11.5hpf expression peak (Fig.7D). Interestingly, the elimination of either the three distal sites in the 'i' region or mutation of the single site in the 5′UTR has only a small (or no) effect on reporter activity (Fig.7D). These results show that multiple *alx1* target sites are required for the normal level of early activation.

However, none of these four P3 sites are responsible for the apparent auto-repression at 16hpf, since the 4-fold mutant shows the same expression profile as the wt construct, albeit with lowered levels of activity (Supplemental Fig.6A). It is this *alx1*-dependant repression that causes the precipitous decline in *alx1* transcript levels after 11.5hpf. Thus when co-injected with *alx1* MASO, construct i' →J is upregulated 2-3 fold at 16 hpf relative to controls, and endogenous *alx1* is up-regulated 3-4 fold (Ettensohn et al., 2003; Supplemental Figure 6B). We made a systematic attempt to scan for sequences responsible for *alx1*-dependant repression, testing a series of deletion constructs of the i' →J reporter. These deletions removed consecutive 100-200bp sequences from the I' -J construct. We calculated the ratio of expression of these reporters at 11.5 to 16 hpf to look for deletion constructs that would exhibit failure of repression (i.e., which would display a ratio closer to 1). However, all constructs showed strong repression at 16hpf relative to 11 hpf (ratios >2.5; Supplementary figure 7). These results indicate that while the auto-activation discovered here is direct, the apparent auto-repression could be either direct or indirect, but in any case the target sites responsible for turning down transcriptional output must not be confined to any one 100-200bp region.

Alx1 *as both activator and repressor*

The evidence points clearly to the janus-like behavior of the Alx1 transcription factor: it functions both as a repressor and an activator (on many down-stream genes as well as itself; Ettensohn et al, 2003; Oliveri et al, 2008). One possibility is that at lower levels of expression, Alx1 exists predominantly as a monomer, which acts as a transcriptional activator, whereas at high levels it dimerizes and becomes a repressor (or attracts a dimer-dependant co-factor which acts as a repressor). To test this, we synthetically generated an obligate dimer form of Alx1 by joining two Alx1 coding sequences with a linker encoding multiples of 4 glycines followed by a serine. Glycine-based linkers are commonly used for this purpose because they lack a β-carbon and therefore permit greater polypeptide flexibility. A serine interspersed between the glycine repeats acts to slow unfolding, thereby providing a useful amount of rigidity to the tether (Robinson and Sauer, 1998). The assumptions used to estimate the minimal length required to join two Alx1 monomers were: 1) that the monomers would be connected from N-terminus to C-terminus (see Fig.8A); 2) that Alx1 can be considered a spherical, globular protein with average density of 0.73cm^3/gm (Harpaz et al., 1994); 3) that the minimum tether distance should at least span the diameter of Alx1; and 4) that the N- and C-termini are not buried within the protein. Under these assumptions, the 440aa Alx1 protein would have a molecular weight of approximately 50kDa and a diameter of 48Å. Given an average peptide unit length of 3.8Å (Iwakura and Nakamura, 1998), we chose a linker sequence repeated 3 times (G4Sx3) to obtain a tether with total length of 57Å. This tether length should be sufficiently long to permit dimerization on antiparallel strands under most protein configurations, except for cases where one monomer's N-terminus is very far from the other monomer's C-terminus. We then analyzed the N- and C- terminal regions of Alx1 and found they contain several hydrophilic residues, suggesting that the ends of Alx1 are indeed exposed, and therefore should be able to be tethered without significantly impairing tertiary structure. The tethered fusion protein was named Alx1-G4Sx3-Alx1. mRNA encoding this protein was injected in a titration experiment into fertilized eggs, and endogenous *alx1* transcript was measured at 11.5 hpf by QPCR. The results (Figure 8B) showed unequivocally that the obligate dimer represses *alx1* by over 2-fold compared to the monomer. Interestingly, in addition to acting as a repressor of *alx1* expression, the Alx1-G4Sx3-Alx1 dimer is a more potent activator of the differentiation *alx1* target gene *msp130L* (Supplemental Figure 8). Taken together, these results show that Alx1 dimerizes to perform auto-regulatory functions of both polarities, but while the P3 palindromic double half-site promotes dimerization on the DNA, higher concentration, for which the obligate tether provides a surrogate, may result in spontaneous formation of dimers that have repressive activity on the alx1 gene itself.

Discussion

Using recombinant BAC reporter knock-ins, and short regulatory expression constructs derived from the BAC, we have solved the *cis*-regulatory system responsible for all aspects of pre-gastrular *alx1* expression, spatial, quantitative, and temporal. Target binding sites and their inputs have been identified. This work has addressed several issues that until now were outstanding: first, what are the specific genomic regulatory features that link the *alx1* gene into the SM double-negative gate control system; second, what are the activators and genomic regulatory features responsible for driving *alx1* expression in the large micromeres; third, what regulatory controls explain the "peak and valley" temporal kinetics of *alx1* message. An outcome of this study is a further elaboration of the GRN subcircuit wiring surrounding the SM double negative gate, as summarized in both the SM lineage and in the rest of the embryo (Figs.9A-B).

Kinetics of alx1 *expression*

As shown in Fig.9C, the temporal *alx1* expression data can be fit in a simple manner on the basis of the conclusions drawn in this work. The assumptions that were used to generate the kinetics shown in Fig.9C are as follows: (1) There is an initial rate of gene expression which obtains from the activation of the gene at about 7.5hpf until enough time has elapsed for the *alx1* mRNA to accumulate and be translated to effective levels (expected to be 2-3h; Bolouri and Davidson, 2003), here taken as until 10hpf. (2) Thereupon a sharp increase in the synthesis rate occurs, due to auto-activation, which from the time-course accumulation data of Fig.1, and the mutation data of Fig.7D, is over twice the initial rate, and this enhanced rate obtains until ~11hpf. (3) The peak of expression is due to the transformation of the Alx1 gene product into a repressor. None of the target sites for Alx1 dimers in constructs displaying this repression mediate repression, and the repressive effect, whether direct or indirect, appears to require distributed sites yet undefined (Supplementary Figures 6A,7A). If the Alx1 dimer binds directly it does so at different sites than those used for activation, but we cannot exclude that it works by activating an unknown repressor which interacts with the DNA at other sites. After the accumulation peak the effect of the repression is to decrease the rate of synthesis, which falls to a steady state obtaining to the end of the period considered, here 17hpf (5). The turnover rate is intrinsic to the mRNA and is constant throughout. It can be estimated from the declining phase of the expression time course, and considering the trough at ~16hpf as a steady state, the synthesis rate after the peak can then be calculated. The parameters used and the kinetic model are shown in the inset in Fig.9C. The point to be made here is that the processes defined experimentally in this work suffice to explain the time course of *alx1* expression. Their significance is straightforward. The auto-activation mechanism serves to drive up the transcript concentration much more rapidly than would otherwise be possible (compare for example the relatively leisurely accumulation of *tbrain* mRNA as shown by Wahl et al; *tbr* is also a target of the double negative gate, but lacks the auto-activation device). But every positive feedback needs to be damped sooner or later or the product accumulates exponentially. The conversion of Alx1 to a repressor is suggested by the synthetic experiment of Fig.8 to be due to a concentration-dependent dimerization mechanism, and the consequence is to self-limit the auto-activation. The decline in *alx1* transcription rate results eventually in a new, lower, steady state preceding the late phase of increased transcription, which is not considered in this paper.

Control of alx1 *expression by the double negative gate*

Functional *cis*-regulatory evidence now directly substantiates the double negative gate regulatory architecture for three target genes, *tbr, delta*, and now *alx1*; and *ets1* probably operates by means of the same types of HesC target sites as have been demonstrated to control spatial expression of these three genes. Of the regulators constituting the definitive SM regulatory state, this leaves only *tel* yet to be examined at the *cis*-regulatory level for evidence that it is a direct double negative gate target gene. The evidence is of the same form for *tbr, delta*, and *alx1*: the predicted HesC target sites are found to be present, and when mutated in expression constructs are demonstrated to be absolutely required to confine expression to the SM domain, just as predicted from extensive earlier data (Oliveri et al., 2002; 2003; Revilla et al, 2007; Oliveri et al, 2008; Smith et al, 2008). The evidence is perhaps strongest for *alx1* and *tbr*, as in both cases the complete genomic landscape in which the gene is embedded, carried in a large BAC recombinant, was subjected to functional *cis*-regulatory analysis, so there is little possibility that a missing regulatory module might have escaped attention. Thus mutating only the two 6 bp HesC target sites of the *alx1* gene in the 129,000 bp BAC produces dramatic ectopic expression. Sharma and Ettensohn (2010) suggested that there is another, different spatial control system confining *alx1* expression to the skeletogenic lineages very early in development, on the basis that when *alx1* is activated at 6th cleavage *hesC* mRNA is also present. However, direct

quantitative measurements both by QPCR and Nanostring nCounter (Revilla et al., 2007; Materna et al, 2010) demonstrate only 1/5th to 1/6th the amount of hesC mRNA at 8hr as at 12hr, amounting to only a few *hesC* mRNA molecules per micromere. The kinetics of *alx1* expression add to the picture when taken together with the *cis*-regulatory evidence that the driver of *alx1* gene expression is Ets1. The Ets1 input indicated in the present experiments conforms perfectly to the conclusions also drawn from *ets1* MASO studies (Ettensohn et al, 2003; Oliveri et al., 2008; Sharma and Ettensohn, 2010). In *S. purpuratus* the *alx1* gene does not become active immediately after the micromeres are born, when the first cohort of SM-specific genes are activated, i.e., *pmar1* and *blimp1* (Oliveri et al, 2002; Revilla et al, 2007; Smith and Davidson, 2008), but only about 2h later, at 7.5-8hpf (Revilla et al, 2007; this work); the useable Ets1 in the micromeres is evidently largely the zygotic product of the newly activated *ets1* gene (see below), and both *ets1* and *alx1* require *pmar1* to have been expressed in order for them to be transcribed normally (*op. cit*.). The experiments in Fig.5C and Fig.S9 show that HesC control of *alx1* transcription extends back to the earliest times its transcript can be reliably measured, only 1.5 to 2h after transcription is initiated during the 6th cleavage cycle, i.e., in the micromeres, at 7th cleavage (Davidson, 1986). It is most unlikely that any earlier spatial control system can be operating. We examined the possibility that the localized maternal regulator Otx, which is a required driver of *pmar1*, could also provide an input into *alx1* even though this would not be consistent with the timing of *alx1* activation, but found that mutation of all the possible Otx target sites in our expression constructs has no effect whatsoever on expression levels (data not shown). This is in contrast with results of just such mutation experiments on the *pmar1* and *blimp1* genes, which are indeed controlled by Otx drivers (Smith and Davidson, 2008; Smith et al, 2007). In summary, *cis*-regulatory evidence surrounding the SM double negative gate now extends from the *pmar1* and *hesC* genes (Smith and Davidson, 2008) to the *tbr* (Wahl et al, 2009), *delta* (Revilla et al, 2007; Smith et al, 2008), and *alx1* genes immediately downstream, and as a result of this study both the spatial and temporal particularities of *alx1* expression have now been incorporated in a consistent explanatory framework based ultimately in genomic regulatory sequence design.

The elegance of the SM double negative gate architecture

In early development spatial expression is controlled at least as much by spatially confined repressors, coupled with wide spread activators, as by locally confined activators (for reviews, Davidson, 2001; 2006). The SM double negative gate is an especially elegant device for initiating spatially confined embryonic gene expression. Its parts list includes the primordially localized micromere inputs Tcf/βcatenin and Otx (Oliveri et al, 2008); one or two zygotically expressed ubiquitous transcriptional activators (unknown); the first zygotically activated, spatially confined positive regulatory gene in the system, *ets1*; a zygotically activated, transcriptionally confined repressor, *pmar1*; a zygotically activated but ubiquitously expressed second repressor, *hesC*; and a set of downstream SM regulatory state genes of which we are here concerned mainly with *alx1*, *tbr*, and *delta*. As discussed by Peter and Davidson (2009), the double negative gate operates globally in a Boolean fashion, in that it not only causes expression of its downstream targets in the confined SM domain, but accomplishes active repression of the same genes everywhere else in the embryo. The discovery that Ets1 is the direct positive driver of *alx1*, just as it is of *delta* and *tbr*, while the *ets1* gene itself is also subject to HesC repression and a target of the double negative gate, adds a beautiful wiring feature to the regulatory architecture. This is captured in Figs.9AB. Here we see, in “View from the Nucleus” BioTapestry models, the sequence of events occurring in the large micromere and its descendants. First, *pmar1* is specifically activated in response to the primordially localized inputs. Then 1.5-2h later, an unknown activator that is evidently ubiquitous appears in the embryo, and turns on the *hesC* gene, everywhere except in the SM lineage, where the *hesC* gene is dominantly repressed by the *pmar1* gene

product (for high resolution relative expression kinetics, see Revilla et al, 2007; Materna et al, 2010). Driven by perhaps the same ubiquitous activator, the *ets1* gene is also turned on, but in space the exact Boolean opposite of *hesC*, only in SM cells, and not elsewhere because the HesC repressor is now elsewhere. The SM-specific, zygotically expressed Ets1 now serves as the driver of the other target genes, *alx1*, *tbr*, and *delta*. This timing, and indeed the fact that the *hesC* double negative gate is what determines SM-specific expression of all these target genes probably means that (the globally distributed) maternally encoded *ets1* mRNA is not the major driver of these genes, or else they would be activated all over the embryo, before the *hesC* gene is activated. Some minor effect of the maternally encoded Ets factor could however account for the residual expression levels of *cis*-regulatory constructs in *ets1* MASO experiments. When the spatial performance of the double negative gate is destroyed, by *hesC* MASO, or by global expression of *pmar1*, global expression of the other target genes results (Oliveri et al., 2002; 2003; Revilla et al., 2007): the reason is that now zygotic expression of their *ets1* driver becomes global. In view of the foregoing we feel the ectopic expression of the Hesc binding site mutant BAC is due not to maternally encoded Ets1/2 in the remainder of the embryo but either to another Ets class factor that binds the Ets target sites (like Ets4) or to an ubiquitous activator that gives a small amount of activity beyond the inputs described in this work. Finally, the relative timing of the successive states of the double-negative gate is a crucial aspect of the spatial control mechanism mediated by the hard wired genomic regulatory circuitry shown in Figs. 9A,B.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr. Joshua Klein for generously providing the G4Sx3 tethering construct. We gratefully acknowledge support for this work from NIH Grant HD-037105 and from The Lucille P. Markey Charitable Trust.

References

Ben-Tabou de-Leon S, Davidson EH. Modeling the dynamics of transcriptional gene regulatory networks for animal development. Dev Biol. 2009; 325:317–28. [PubMed: 19028486]

Beverdam A, Meijlink F. Expression patterns of group-I aristaless-related genes during craniofacial and limb development. Mech Dev. 2001; 107:163–7. [PubMed: 11520673]

Brown CT, Rust AG, Clarke PJC, Pan Z, Schilstra MJ, De Buysscher T, Griffin G, Wold BJ, Cameron RA, Davidson EH, Bolouri H. New computational approaches for analysis of cis-regulatory networks. Developmental Biology. 2002; 246:86–102. [PubMed: 12027436]

Cheers MS, Ettensohn CA. Rapid microinjection of fertilized eggs. Methods Cell Biol. 2004; 74:287–310. [PubMed: 15575612]

Court DL, Sawitzke JA, Thomason LC. Genetic engineering using homologous recombination. Annu. Rev. Genet. 2002; 36:361–88. [PubMed: 12429697]

Davidson, EH. Gene Activity in Early Development. Academic press; Orlando: 1986.

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C-H, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Schilstra MJ, Clarke PJC, Rust AG, Pan Z, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Developmental Biology. 2002; 246:162–90. [PubMed: 12027441]

Ettensohn CA, Illies MR, Oliveri P, De Jong DL. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling

skeletogenic fate specification in the sea urchin embryo. Development. 2003; 130:2917–28. [PubMed: 12756175]

Fischer A, Gessler M. Delta-Notch--and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. Nucleic Acids Res. 2007; 35:4583–96. [PubMed: 17586813]

Gao F, Davidson EH. Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. Proc Natl Acad Sci USA. 2008; 105:6091–6. [PubMed: 18413604]

Grbavec D, Stifani S. Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. Biochem Biophys Res Commun. 1996; 223:701–5. [PubMed: 8687460]

Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. Structure. 1994; 2:641–9. [PubMed: 7922041]

Hobert O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. BioTechniques. 2002; 32:728–30. [PubMed: 11962590]

Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Dev Biol. 2006; 300:90–107. [PubMed: 17054934]

Iwakura M, Nakamura T. Effects of the length of a glycine linker connecting the N-and C-termini of a circularly permuted dihydrofolate reductase. Protein Eng. 1998; 11:707–13. [PubMed: 9749924]

Lee PY, Nam J, Davidson EH. Exclusive developmental functions of gatae cis-regulatory modules in the Strongylocentrorus purpuratus embryo. Developmental Biology. 2007; 307:434–45. [PubMed: 17570356]

Materna SC, Nam J, Davidson EH. High accuracy, high-resolution prevalence measurement for the majority of locally expressed regulatory genes in early sea urchin development. Gene Expr Patterns. 2010; 10:177–184. [PubMed: 20398801]

Nam J, Dong P, Tarpine R, Istrail S, Davidson EH. Functional cis-regulatory genomics for systems biology. Proc Natl Acad Sci USA. 2010; 107:3930–5. [PubMed: 20142491]

Oliveri P, Carrick DM, Davidson EH. A regulatory gene network that directs micromere specification in the sea urchin embryo. Developmental Biology. 2002; 246:209–28. [PubMed: 12027443]

Oliveri P, Davidson EH, McClay DR. Activation of pmar1 controls specification of micromeres in the sea urchin embryo. Developmental Biology. 2003; 258:32–43. [PubMed: 12781680]

Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic cell lineage. Proc Natl Acad Sci USA. 2008; 105:5955–62. [PubMed: 18413610]

Paroush Z, Finley RL Jr. Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. Cell. 1994; 79:805–15. [PubMed: 8001118]

Qu S, Li L, Wisdom R. Alx-4: cDNA cloning and characterization of a novel paired-type homeodomain protein. Gene. 1997; 203:217–23. [PubMed: 9426253]

Revilla-i-Domingo R, Minokawa T, Davidson EH. R11: a cis-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres. Developmental Biology. 2004; 274:438–51. [PubMed: 15385170]

Revilla-i-Domingo R, Oliveri P, Davidson EH. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. Proc Natl Acad Sci USA. 2007; 104:12383–8. [PubMed: 17636127]

Rizzo F, Fernandez-Serra M, Squarzoni P, Archimandritis A, Arnone MI. Identification and developmental expression of the ets gene family in the sea urchin (Strongylocentrotus purpuratus). Dev Biol. 2006; 300:35–48. [PubMed: 16997294]

Robinson CR, Sauer RT. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proc Natl Acad Sci U S A. 1998; 95:5929–34. [PubMed: 9600894]

Sharma T, Ettensohn CA. Activation of the skeletogenic gene regulatory network in the early sea urchin embryo. Development. 2010; 137:1149–57. [PubMed: 20181745]

Smith J, Davidson EH. Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo. Proc Natl Acad Sci USA. 2008; 105:20089–94. [PubMed: 19104065]

Wahl M, Hahn J, Gora K, Davidson E, Oliveri P. The cis-regulatory system of the tbrain gene: Alternative use of multiple modules to promote skeletogenic expression in the sea urchin embryo. Developmental Biology. 2009

Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG. Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res. 2005; 33:e36. [PubMed: 15731329]

Wilson D, Sheng G, Lecuit T, Dostatni N, Desplan C. Cooperative dimerization of paired class homeo domains on DNA. Genes & Development. 1993; 7:2120–34. [PubMed: 7901121]

Yajima M, Umeda R, Fuchikami T, Kataoka M, Sakamoto N, Yamamoto T, Akasaka K. Implication of HpEts in gene regulatory networks responsible for specificaiton of sea urchin skeletogenic primary mesenchyme cells. Zoological Science. 2010; 8:638–46. [PubMed: 20695779]

Research Highlights

identification of promoter-proximal region that regulates alx1 expression in the embryonic skeletogenic precursors of sea urchin, *Strongylocentrotus purpuratus*

identification of hesc, ets1, and alx1 itself as regulators of alx1

use of high-throughput tagged reporter system for cis-regulatory analysis

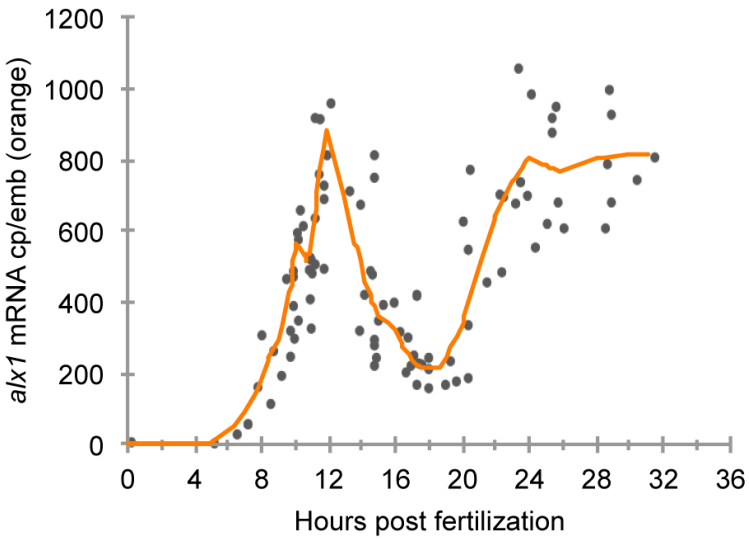

Fig 1. High-density timecourse of endogenous *alx1* expression

Measurements of *alx1* mRNA abundance were compiled from multiple (n=8) experiments over the course of the first 18 hours of development and smoothed by LOWESS regression (orange line).

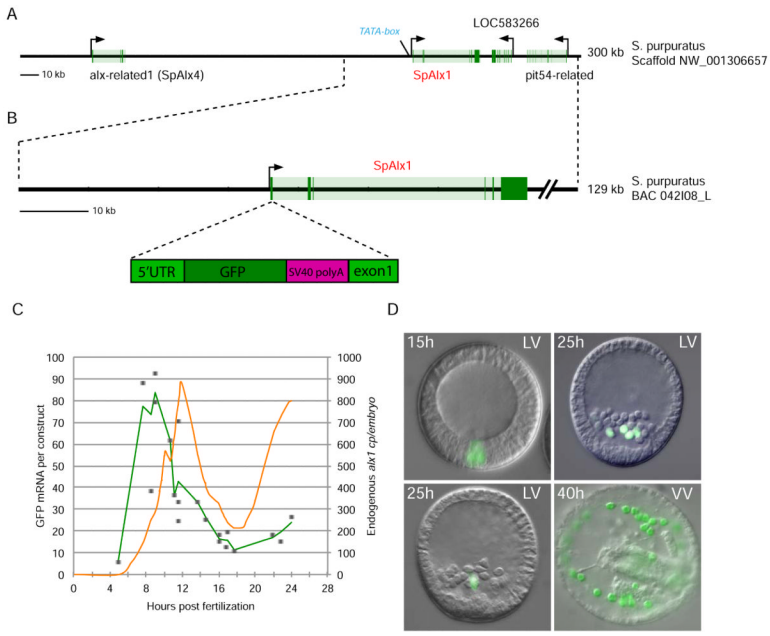

Fig 2. Correct spatiotemporal expression of alx-GFP BAC reporter construct

A) Scaffold NW_001306657 contains 4 genes including *alx1*. The alx1-related gene, *sp-alx4* (NCBI Ref Seq. XM_780145, Baylor Gene GLEAN3_22816) lies roughly 160kb upstream of *alx1* and is oriented in the same direction. Two genes lie within 50kb downstream of *alx1* and are oriented opposite to *alx1*, LOC583266 and *pit54-related* (NCBI Ref Seq. XM_783163.2). A canonical TATA box sits at −49 bp (blue). Exons, dark-green; introns, light-green; arrowhead, transcriptional start site **B)** A 140kb BAC (Sp_042I08_L) isolated from an *S.p*. genomic library was found to contain *alx1*. The coding sequence is flanked upstream by 35kb of genomic DNA and downstream by 65kb. A cassette containing GFP coding sequence and SV40polyA 5′UTR was inserted at the start of *alx1* coding sequence as described **C)** GFP mRNA expression in embryos injected with the Alx-GFP BAC was measured by qPCR from 5-24hpf. Data was combined from several experiments (n=5) and smoothed using LOWESS regression (green line). Endogenous *Alx1* expression timecourse is overlayed on the secondary axis (orange line). **D)** BAC-GFP injected embryos were imaged for GFP fluorescence and overlayed onto DIC pictures at 15, 25 and 40hpf (top row 15 and 25hpf, bottom row 25 and 40hpf). LV, lateral view; VV, ventral view.

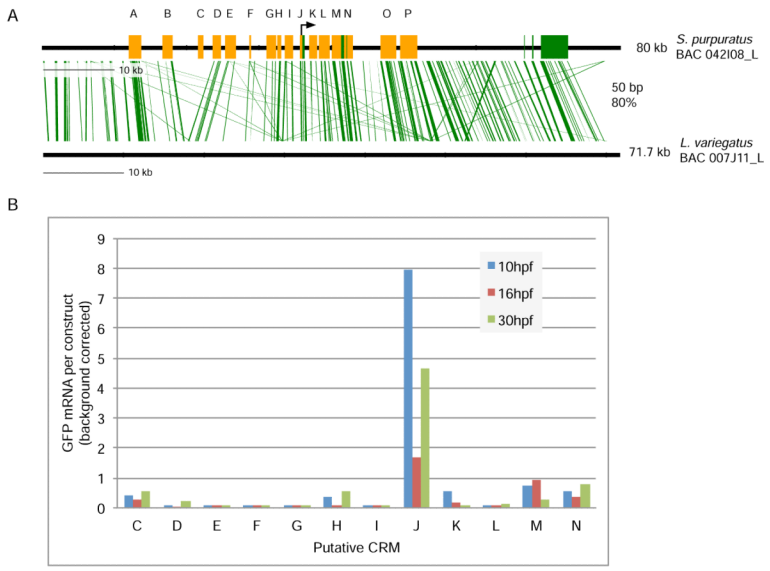

Fig 3. Phylogenetic footprinting and scanning of activity of cis-elements lying within 15kb of the *alx1* promoter

A) A 150kb BAC (Lv_007J11_L) was isolated from an *L.v.* genomic library and sequenced. Phylogenetic footprinting was performed using seqComp (Brown et al., 2002) with a 50bp/80% identity window. 14 conserved elements (labeled A-N) were found to lie between −22kb and +8kb relative to the first exon. **B)** 12 elements were examined for their ability to drive expression at 10,16 and 30hpf when fused to a GFP expression cassette containing the *alx1* basal promoter. Multiple constructs were measured independently using a tagged GFP technology and normalized for differences in the number of integrated genomic copies. Expression data was corrected for background expression as described in materials and methods.

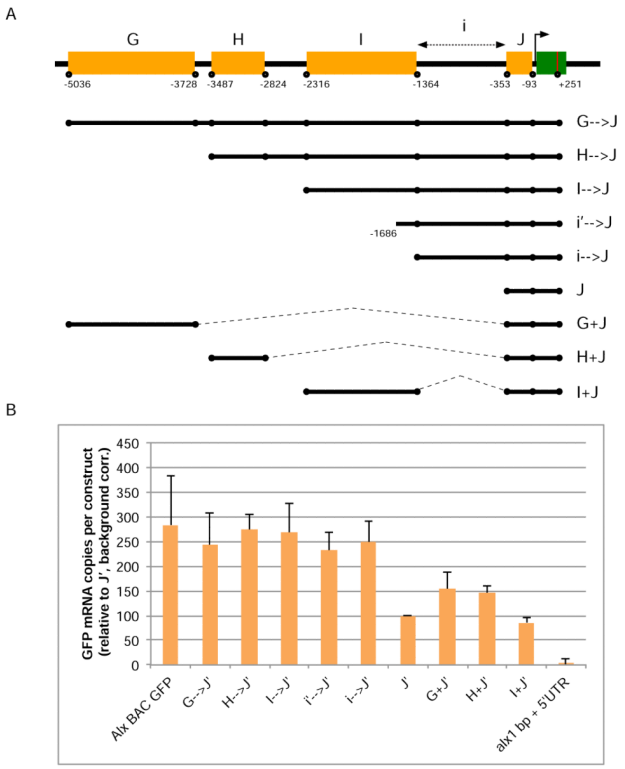

Fig 4. Serial Truncation of Alx1 GFP BAC reporter construct identifies a non-conserved, functional regulatory sequence

A) A series of tagged-GFP reporter constructs were generated to test the activity of regulatory sequences upstream of the promoter-proximal conserved module J **B)** qPCR data of injected reporter constructs was compared against activity of the full-length *alx1* GFP BAC at 11-12hpf. Construct i' →J contains a conserved subregion of module I but is identical in expression to construct i→J. Data are represented as normalized GFP expression relative to a reporter construct containing the J-module, *alx1* basal promoter and *alx1* 5′UTR. Expression data was also normalized for tag-specific variation as described. Expression data was corrected for background expression as described in materials and methods.

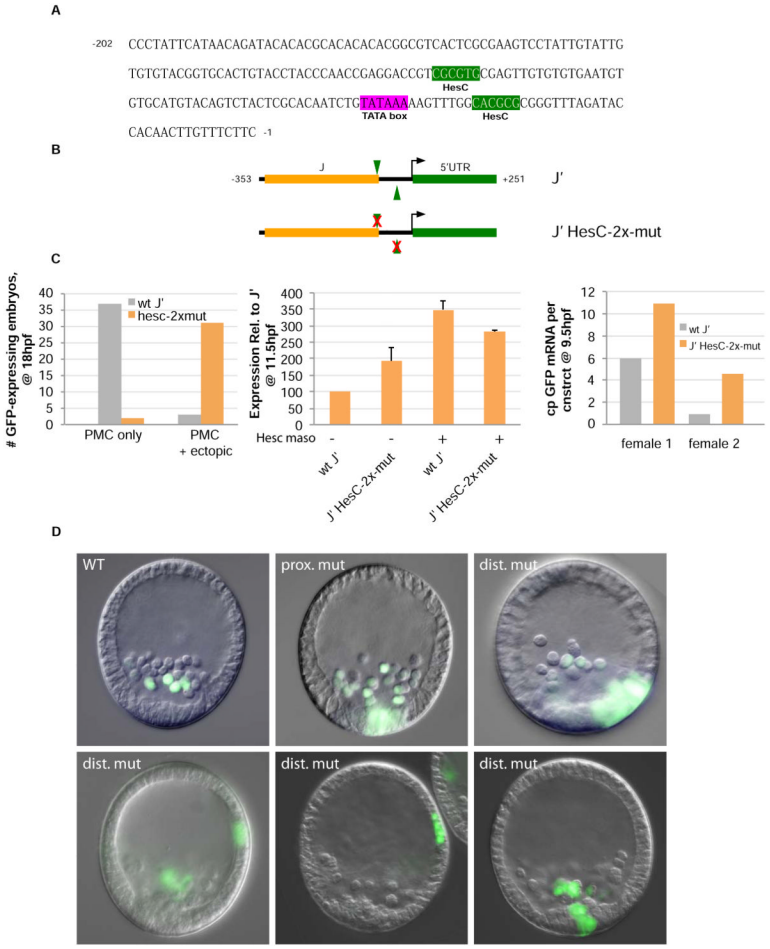

Fig 5. Effects of mutation of promoter-flanking Class-C bHLH sites on *alx1* expression

A) Sequence of the promoter-proximal region, showing two class-C bHLH binding sites (green highlight) flanking the TATAA box (purple highlight). **B)** Map of J construct, showing conserved J-module (orange), *alx1* 5′UTR (green) and class-C bHLH sites (green triangles) in antiparallel orientation (top) and map of mutant J construct where both bHLH binding sites are mutated as described. **C)** Left, spatial expression pattern of JHesC-2x-mut relative to wild type J at 11.5hpf. Data reported is the number of embryos out of 100. Middle, QPCR assay of reporter GFP expression of HesC binding site mutant construct with and without *hesC* MASO. Expression is reported relative to wild-type J construct shown in (B) and data are normalized for number of incorporated constructs and tag-specific variation (n=3) Expression data was corrected for background expression as described in materials and methods. Right, reporter activity at 9.5hpf of wild-type J construct and HesC-2x-mut. Data are reporter as copies GFP mRNA per construct and corrected for background expression level. **D)** Spatial expression pattern of BAC-GFP constructs whereby either the downstream (prox. mut) or upstream (dist. mut) HesC binding site is mutated. Top left, wild-type BAC; top middle, proximal-HesC site mutant BAC; top right and bottom row, distal-HesC site mutant BAC.

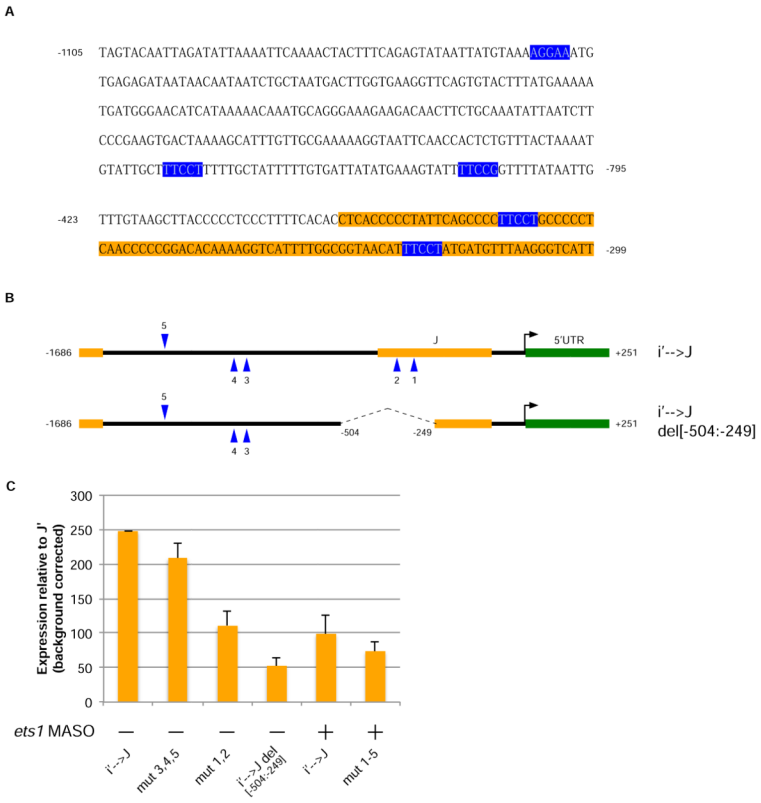

Fig 6. Ets1 sites in i→J are necessary for expression

A) Sequences within module i and J that contain Ets binding sites (blue highlight) of the form MGGAA. J-module sequence is highlighted in yellow. **B)** Map of i→J construct and deletion construct showing Ets1 binding sites (blue triangles). **C)** QPCR assay of reporter GFP activity of Ets1-binding site mutant and deletion constructs in the presence of Ets1 MASO at 11-12hpf. Activity is reported relative to expression of wt construct i→J shown in (B) and data are normalized for number of incorporated constructs and tag-specific variation (n=4). Expression data was corrected for background expression as described in materials and methods.

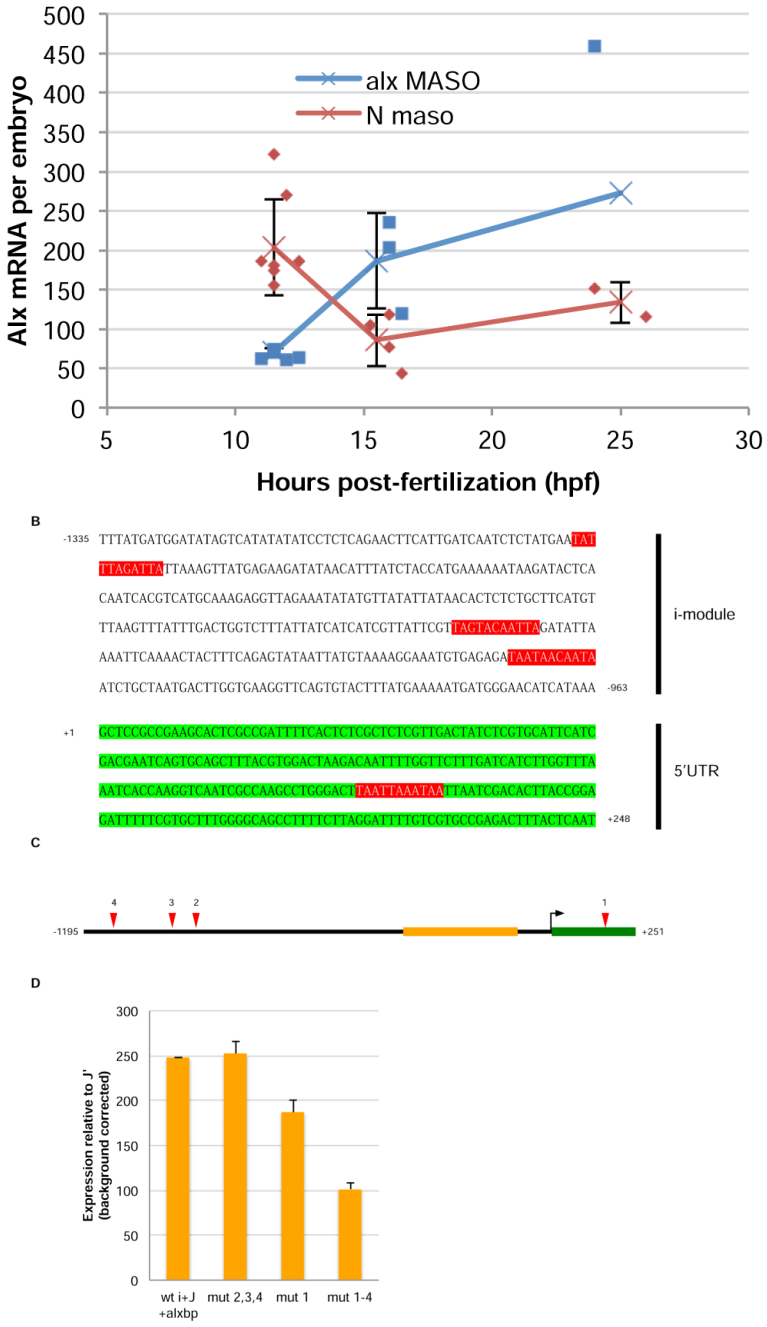

Fig 7. P3 sites in modules i and J drive *alx1*

A) qPCR timecourse of endogenous *alx1* mRNA in response to alx1 MASO (blue line) and control MASO (red line) injection. **B)** Sequences within modules i and J that contain Alx1/Cart family binding sites of the form TAATNNNATTA. **C)** Map of i→J construct showing Alx1 p3 sites (red triangles). **D)** qPCR assay of reporter GFP activity of Alx1-binding site mutant constructs at 11-12hpf. Activity is reported relative to wild-type i→J expression and data are normalized for number of incorporated constructs and tag-specific variation (N=3). Expression data was corrected for background expression as described in materials and methods.

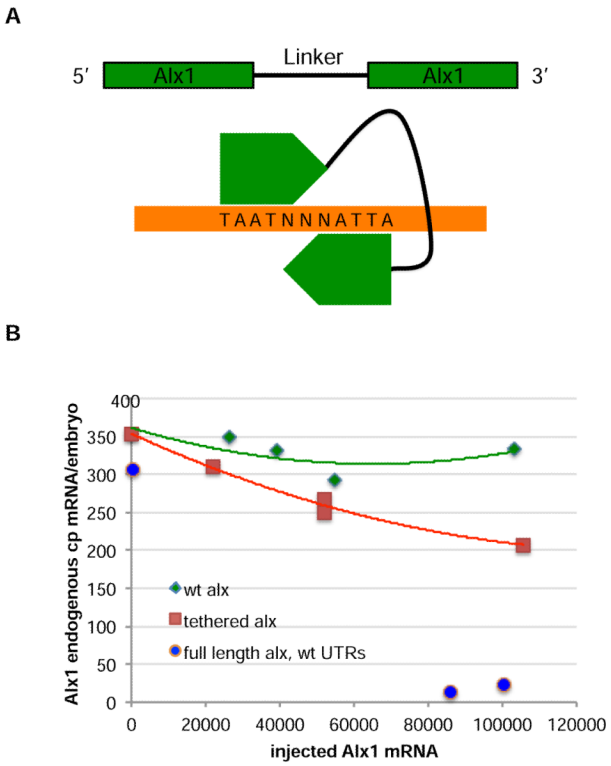

Fig 8. Alx1 MASO has dual effects on *alx1* transcription

A) (top) An obligate dimer form of Alx1 was generated by inserting sequence encoding a G4Sx3 (glycinex4 + serine) tether in between two *alx1* coding sequences. This construct was fused to endogenous 5′UTR and 3′UTR sequence and in-vitro transcribed and injected into fertilized embryos. (bottom) A model showing Alx1 homodimer binding in an antiparallel orientation to target P3 sites. Linker length is estimated at 57 Angstroms and diameter of a single Alx1 monomer is estimated to be roughly 48 Angstroms. **B)** Effect on endogenous *alx1* mRNA levels in response to a titration of *alx1* mRNA or *tethered alx1* mRNA injection. Abscissa is injected mRNA levels measured at 11-12hpf and ordinate reflects levels of endogenous *alx1* mRNA. Injection of wild-type *alx1 mRNA* containing endogenous alx1 5′UTR and 3′UTR improved the efficiency of repression by > 10 fold.

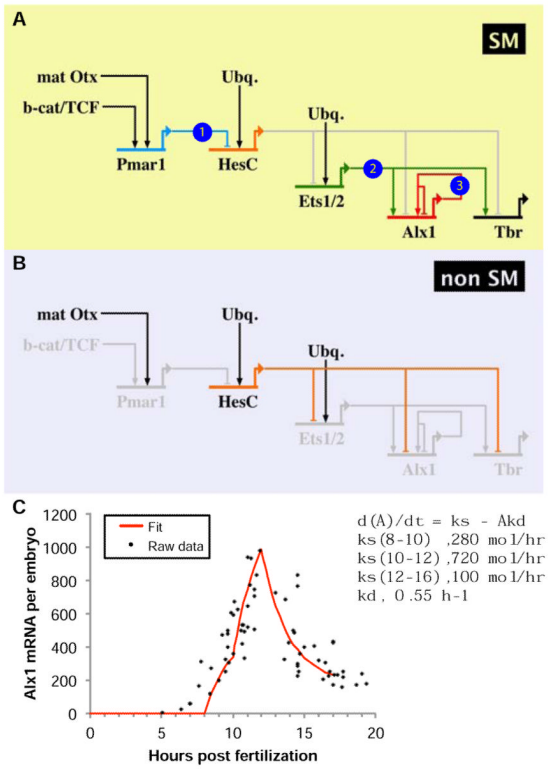

Fig 9. Network architecture controlling *alx1* expression in large micromeres and resultant kinetics

A) A view from the nucleus of large micromeres and their descendents. Inputs are labeled in order of the relative order of their effects on alx1 transcription. 1) As a result of initial expression of *pmar1* at the end of 4th cleavage, *hesc* is cleared from the large micromeres, 2) at 8-10 hpf *ets1/2* initiates *alx1* transcription, 3) from 10-12 hpf Alx1 protein autoactivates *alx1* and at 12 hpf Alx1 begins to function as an autorepressor. **B)** View from the nucleus of nonskeletogenic cell types (the rest of the embryo) through mid-blastula stage. *hesc* continues to be expressed in these cells and blocks transcription of *ets1/2*, which prevents the expression of *alx1* and *tbrain* as well. **C)** The kinetics of *alx1* transcription from Figure 1 were fit to the differential equation $d(A)/dt = k_s - Ak_d$ (Ben-Tabou de-Leon and Davidson, 2009) using the following parameter values: k_d is 0.55 h^{-1}, and k_s is the synthesis rate which varies as follows: from 8-10 hpf, k_s = 280 molecules per hour (mol/hr); from 10-12 hpf, k_s =720 mol/hr; and from 12-17 hp, k_s=100 mol/hr. These transcriptional rates are well within the limits of sea urchin transcriptional machinery at 15°C. Considering the most rapid rate of *alx1* transcription, 720 molecules/hr, and given there are 8 large micromeres containing 16 copies of the *alx1* gene, the initiation rate per gene copy is 45 transcripts/hr or 1 transcript every 80 seconds. This value is roughly 8 times slower than the computed maximal initiation rate at 15C of 1 transcript every 9 seconds (Davidson, 1986; pp.144-145)

Table1

Expression of GFP in embryos injected with reporter constructs

Construct @ 18-24 h	**No. Embryos observed**	**No. GFP + embryos** *	**% SM**	**%NSM**	**%Endoderm**	**% Ectoderm**
Alx GFP BAC	951	434(46)	369(85) 330(76)**	44(10)	10(2)	54(12)
Alx GFP BAC, prox Hesc mut	125	86(69)	67(78)	37(43)	0(0)	13(15)
Alx GFP BAC, dist. HesC mut	147	76(52)	62(82)	19(25)	8(11)	22(29)
GHIJ-GFP construct	158	60(38)	45(75)	13(22)	0(0)	5(8)
J'-construct	80	45(56)	40(89)	5(11)	0(0)	0(0)
Alx bp	100	0	0	0	0	0

* percentage expressing embryos in each category are in parentheses

** number/percentage expressing *only* in PMCs

Dev Biol. Author manuscript; available in PMC 2012 September 15.